Too many ghosts in the machine [Review] Beth Singler (2019) Ghost in the Machine

Article (Accepted Version)

Too many ghosts in the machine

Anil Seth
Sackler Centre for Consciousness Science and Dept of Informatics
University of Sussex
Brighton
BN1 9QJ
a.k.seth@sussex.ac.uk

April 27, 2019

In Alex Garland's 2015 *Ex Machina*, there's this one scene where beautiful-intelligent-robot Ava asks coder-cum-stooge Caleb a simple question: "Do you think about me, when we aren't together?" Caleb is supposed to be figuring out whether Ava is conscious or not, but here Ava turns the tables. She is now the one trying to figure out, and perhaps manipulate, Caleb. And before long all hell breaks loose.

Garland's landmark film is heavily referenced in Beth Singler's *Ghost in the Machine,* made in association with Cambridge University's Leverhulme Centre for the Future of Intelligence. This is no big-budget Hollywood production - mind you neither was *Ex Machina* - but its targets are the same. What is the relationship between intelligence and consciousness? Could a machine, however advanced, have feelings? These are fascinating, controversial, and important questions and Singler covers an awful lot of ground in just fifteen minutes. No less than fifteen people offer all sorts of opinions, interspersed with scenes of a supposedly remarkable child – a weird amalgam of Ava and Eleven from *Stranger Things* – being quizzed about 'what it is like to be her' in transparent homage to *Ex Machina*.

From one perspective Singler's strategy makes good sense. There is such a diversity of views on the topic of 'machine' or 'artificial' consciousness that there's clear value in bringing this intellectual hodgepodge to light and letting the viewer make up their own mind. *Ghost in the Machine* does work well on this level. In the end, though, I was left without a clear sense of direction or purpose, perhaps an inevitable trade-off for a short film that attempts to cover so much.

Let's take the questions one at a time. First up is the relationship between consciousness and intelligence. Does consciousness require intelligence – and does intelligence, perhaps on exceeding some as-yet-unknown threshold, entail consciousness? It's easy to assume that the two go hand-in-hand, but it's equally easy to see that this assumption might be a trivial consequence of an unexamined anthropocentrism: *we* are intelligent (more than the rest) and *we* are conscious (and we don't really know about the rest) so intelligence and consciousness must go together.

Early in Singler's film we hear from Lucy Cheke of Cambridge University's Department of Psychology. Cheke has worked with Orang-utans and Eurasian Jays, and she perceives a flicker of human-like consciousness in both. Then Marta Halina, programme director at the Leverhulme Centre, cuts in to suggest that if we discovered that consciousness underpinned intelligence then we'd have to conclude that intelligence implied consciousness. But before

she can explain what this might mean in practice the film swerves again, this time to an explanation of an EEG experiment about face recognition which, despite the claims of the experimenter, doesn't really connect either with consciousness or with the 'sense of reality' that is a pervasive though often overlooked feature of everyday conscious experience.

This sets the pattern for the rest of the film. Individually, the interview fragments promise much, but collectively they struggle. This is particularly obvious when we meet Father Ezra Sullivan from the Pontifical University of St Thomas Aquinas. In a fascinating snippet, Fr Sullivan explains that the process of human reproduction is more than "the making of a machine because it requires, according to our theological understanding, the intervention of God to create a soul for each human body … there's always a gap there, between robot nature and human nature". Instead of examining what on earth Fr Sullivan might be on about, Singler moves swiftly to the Chairman of the Cambridge Buddhist Centre who offers up a distinction between ordinary consciousness and awakened consciousness before hedging his bets on whether an artificial intelligence – an AI – could ever have feelings.

And through these scenes we've segued surreptitiously from the relationship between intelligence and consciousness to the more sci-fi ready topic of machine consciousness. There are now three properties in play: consciousness, intelligence, and the property of being alive. Might *being alive* be the key requirement for consciousness, rather than being intelligent, or processing information in some particular way? This is where Fr Sullivan's views about the soul invite further analysis. Early conceptions of soul, predating Sullivan's theological understanding, were in fact closely associated with the living body. The Hindu concept of Ātman interpreted soul as 'breath,' and carried no presumption that personal identity would survive death of the body. In the West, the relationship between consciousness, soul, and life was put under the spotlight by the 18th Century polymath Julian Offray de La Mettrie's controversial *L'homme machine* argument [1]. While Descartes had decided that non-human animals were merely 'beast machines', lacking in reason and consciousness (and soul), La Mettrie proposed that if animals were machines made of meat then so too were humans, though he left tantalizingly ambiguous whether in making this argument he was trying to mechanise the soul, or give reason and awareness to the mechanical [2]. Either way the Catholic church didn't much like it. Perhaps not surprisingly, I like it very much and think that approaching the brain from the perspective of performing physiological regulation – as opposed to implementing some kind of substrate-independent intelligence – sheds considerable light on conscious perception, both of the world and of the self [3].

The next voices bring a welcome contemporary perspective. Ron Chrisley, my colleague at the University of Sussex, imagines a robot describing a visual scene in the same way a human would. What if the robot seemed subject to the same kinds of visual illusions as we humans – would that be evidence of robot 'experience'? I'm not convinced, but neither am I convinced that this is what Chrisley intended to convey. Then Murray Shanahan, of Imperial College London and latterly also DeepMind, hits the nail on the head by pointing out that advanced AI might not require consciousness at all, a point echoed a minute later by Anders Sandberg from the Future of Humanity Institute at the University of Oxford. Shanahan also suggests that what we think of as normal human consciousness can be broken down into

several potentially separable components which might manifest in different ways in other species or perhaps in future machines – a view I've also argued for [4].

Prospects of possibly conscious, potentially general, advanced AI raise a host of ethical concerns which speed by in the remainder of the film. As Chrisley remarks, even if true machine consciousness is impossible or very distant, near-future robots might be able to convince us that they have real feelings. Navigating such a landscape with our distinctly unfit-for-purpose moral compasses, tuned to seemingly self-evident distinctions between the living and the non-living, the aware and the unaware, might prove tricky indeed. Hanging over this landscape, as the philosopher Thomas Metzinger has pointed out (though not in this film), is the potential ethical disaster of introducing new forms of machine suffering into the world, a possibility that needs to be taken seriously even if it is very remote [5]. We hear similar sentiments in a rapid-fire parade of interview fragments, including one from Singler herself.

Separating hype from reality in discussions of AI and consciousness is a difficult task. Within AI, exciting developments in what often amounts to statistical pattern recognition can be hard to reconcile with popular conceptions of AI as portrayed in science fiction. And consciousness research, while now firmly re-established at the heart of psychology and neuroscience [6], remains a magnet for fringe opinions and wild speculation. Kudos then to Singler for leading us through a thicket of diverse opinions without force-feeding us any particular perspective. Kudos also for giving the last word (besides Ava/Eleven and her mysterious interlocutor) to Murray Shanahan who reminds that technological developments need not always be threatening, but can also replenish the wonder we take in what we are and in the variety of other minds that surround us. All in all, *Ghost in the Machine* dishes up a serviceable appetizer but for the main course be sure to leave room for *Ex Machina*.

References

1. de La Mettrie, J.O., *L'homme machine*. 1748, Leiden: Luzac.
2. Makari, G., *Soul machine: The invention of the modern mind*. 2016, London: W. W. Norton.
3. Seth, A.K. and M. Tsakiris, *Being a Beast Machine: The Somatic Basis of Selfhood.* Trends Cogn Sci, 2018. **22**(11): p. 969-981.
4. Seth, A.K. *The real problem*. Aeon 2016.
5. Metzinger, T., *Suffering*, in *The return of consciousness: A new science on old questions*, K. Almqvist and A. Haag, Editors. 2017, Axel and Margaret Ax:son Johnson Foundation: Stockholm. p. 221-249.
6. Michel, M., et al., *Opportunities and challenges for a maturing science of consciousness.* Nat Hum Behav, 2019. **3**(2): p. 104-107.